IT & C

ISSN 2821 - 8469, ISSN – L 2821 - 8469, Volumul 1, Numărul 1, Septembrie 2022

Securitatea Big Data: Amenințări

Nicolae Sfetcu

Pentru a cita acest articol: Sfetcu, Nicolae (2022), Securitatea Big Data: Amenințări, *IT & C*, **1**:1, 46-59, DOI: 10.58679/IT72548, https://www.internetmobile.ro/securitatea-big-data-amenintari/

Publicat online: 21.08.2022

Securitatea Big Data: Amenințări

Nicolae Sfetcu

Rezumat

Taxonomia amenințărilor este una cuprinzătoare, cu un accent special pe amenințările de securitate cibernetică; adică amenințări care se aplică activelor tehnologiei informației și comunicațiilor. Au fost considerate amenințări suplimentare care nu derivă din TIC pentru a acoperi amenințările asupra bunurilor fizice și, de asemenea, atât dezastrele naturale (care nu sunt declanșate direct de oameni), cât și dezastrele de mediu cauzate direct de oameni.

Cuvinte cheie: securitate, Big Data, megadate, amenințări

IT & C, Volumul 1, Numărul 1, Septembrie 2022, pp. 46-59
ISSN 2821 - 8469, ISSN – L 2821 - 8469
URL: https://www.internetmobile.ro/securitatea-big-data-amenintari/

Taxonomia amenințărilor

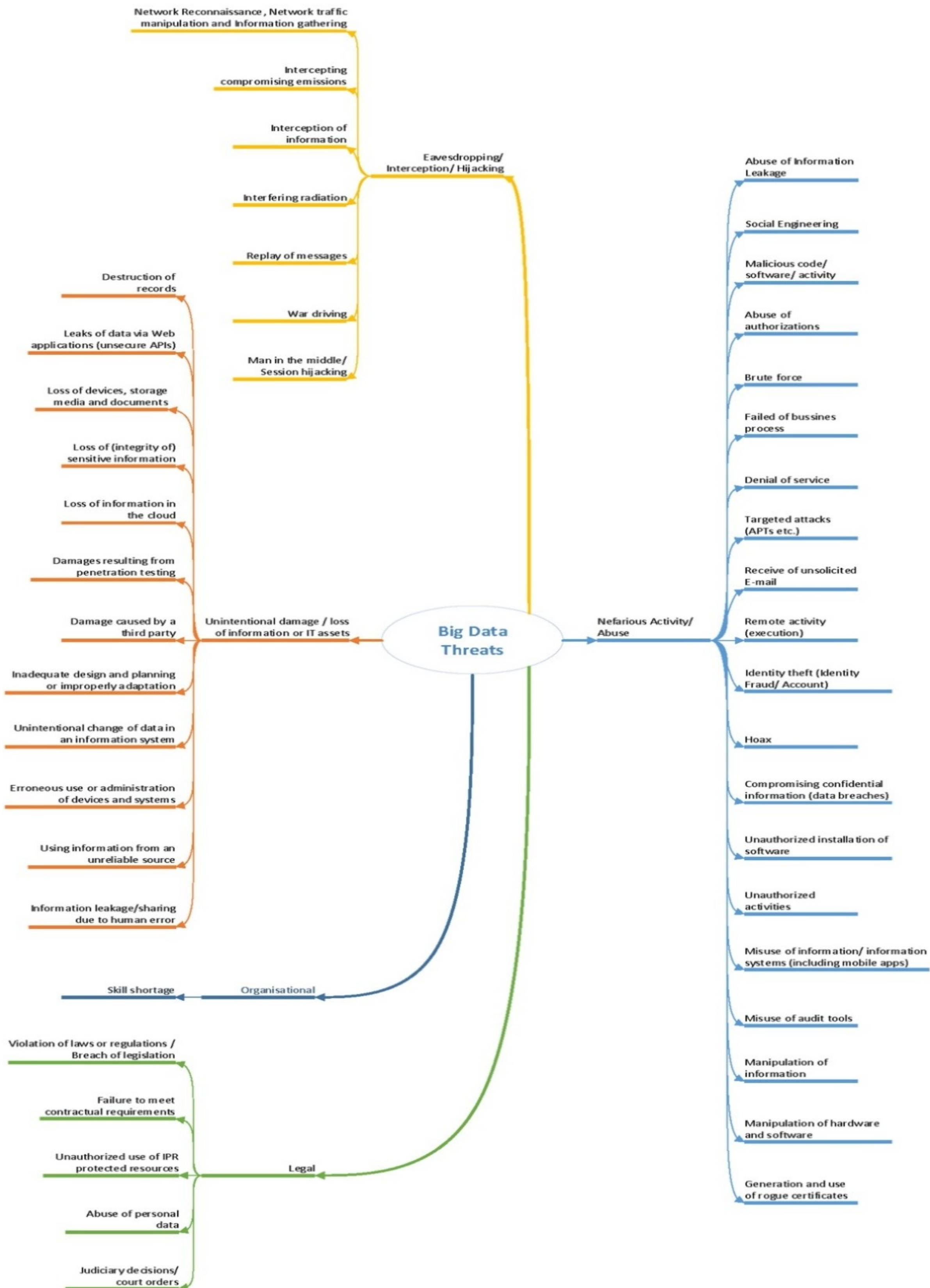
Taxonomia amenințărilor este una cuprinzătoare, cu un accent special pe amenințările de securitate cibernetică; adică amenințări care se aplică activelor tehnologiei informației și comunicațiilor. Au fost considerate amenințări suplimentare care nu derivă din TIC pentru a acoperi amenințările asupra bunurilor fizice și, de asemenea, atât dezastrele naturale (care nu sunt declanșate direct de oameni), cât și dezastrele de mediu cauzate direct de oameni.

Taxonomia amenințărilor a fost dezvoltată de Grupul ENISA Threat Landscape (ETL) și reprezintă o consolidare a amenințărilor luate în considerare anterior în alte rapoarte tematice (1) și cercetări ample. Taxonomia include amenințări aplicabile activelor Big Data și numai acestea sunt descrise în figură. În subsecțiunea următoare, amenințările specifice Big Data care au fost identificate printr-o literatură extinsă, care au fost atribuite categoriilor relevante definite în taxonomia de amenințare a ENISA, sunt mapate la taxonomia activelor Big Data discutate anterior.

Maparea amenințărilor la activele Big Data

Această analiză se bazează pe o analiză extinsă a incidentelor de amenințare reală și a atacurilor la Big Data prezentate în articole, bloguri tehnice, lucrări de conferință, precum și sondaje online pentru colectarea de informații suplimentare. Revizuirea a fost determinată de taxonomia generică a amenințărilor ENISA prezentată în secțiunea anterioară.

În termeni generali, amenințările, precum întreruperea rețelei sau defecțiunile infrastructurii de suport, pot afecta puternic Big Data. De fapt, întrucât un set Big Data are milioane de bucăți de date și fiecare bucată poate fi localizată într-o locație fizică separată, această arhitectură duce la o dependență mai mare de interconectările dintre servere. Rapoartele tematice ENISA anterioare au abordat în profunzime amenințările precum întreruperile și defecțiunile, care afectează legăturile de comunicații în rețea (2). Din acest motiv, aici nu luăm în considerare aceste amenințări. De asemenea, am ales să nu ne oprim asupra atacurilor fizice (deliberate și intenționate), a dezastrelor naturale și de mediu și a eșecurilor / defecțiunilor (de exemplu, disfuncționalități ale infrastructurii de suport TIC), deoarece efectele lor sunt puternic atenuate de redundanța intrinsecă a Big Data, deși proprietarii de Big Data care își instalează sistemele în cloud-uri private sau alte infrastructuri locale ar trebui să ia în considerare aceste atacuri (3).



(*Taxonomie de amenințare aplicabilă activelor Big Data*)

În general, o amenințare este „orice circumstanță sau eveniment cu potențialul de a avea un impact negativ asupra unui activ prin acces neautorizat, distrugere, divulgare, modificare a datelor și / sau refuz de serviciu” (4). Având în vedere definiția pe care am dat-o despre Big Data (volum, viteză, varietate, veridicitate, variabilitate și valoare), o amenințare la adresa unui activ Big Data poate fi considerată ca orice circumstanță sau eveniment care afectează, adesea simultan, volume mari de date și / sau date din diverse surse și de diferite tipuri și / sau date de mare valoare.

De asemenea, identificăm două sub-categorii diferite de amenințări: breșe („Big Data Breach”) și scurgeri („Big Data Leak”) (5), ortogonal cu taxonomia amenințărilor utilizate. O breșă apare atunci când „un activ de informație digitală este furat de atacatori prin pătrunderea în sistemele sau rețelele TIC în care este deținut / transportat” (6). Putem defini „breșa Big Data” ca furtul unui activ Big Data executat prin pătrunderea în infrastructura TIC. O scurgere de Big Data, pe de altă parte, poate fi definită ca divulgarea (totală sau parțială) a unui activ Big Data într-o anumită etapă a ciclului său de viață. O scurgere de date mari se poate întâmpla, de exemplu, în proiectarea inadecvată, adaptarea software necorespunzătoare, sau atunci când un proces de afaceri eșuează. În ceea ce privește modelul atacatorului, o breșă Big Data necesită un comportament ostil pro-activ (spargerea), în timp ce o scurgere Big Data poate fi exploatată chiar de atacatori cinstiți, dar curioși.

Grupul de amenințări: Daune neintenționate/pierderea informațiilor sau a activelor IT

Acest grup de vulnerabilități include scurgerea de informații sau partajarea din cauza erorilor umane, intervenția neintenționată sau utilizarea eronată a administrării sistemelor (configurare greșită), pierderea dispozitivelor.

Amenințare: Surgerea/partajarea informațiilor din cauza unei erori umane

Amenințările accidentale sunt cele care nu sunt favorizate în mod intenționat de oameni. Acestea se datorează configurării greșite, flisărilor neintenționate și erorilor de scris (de exemplu apăsarea butonului greșit), aplicării greșite a regulilor valide (gestiune slabă a corecțiilor, utilizarea numelor și parolelor implicite de utilizator sau a parolelor ușor de ghicit) și greșelilor bazate pe cunoștințe (actualizări de software și blocări, probleme de integrare, defecte procedurale) (7, 8).

Scurgerea de informații din cauza configurării greșite poate fi o problemă comună: conform unui studiu recent 9, configurațiile eronate de administrare a sistemului au condus la numeroase deficiențe în patru tehnologii diferite de Big Data; adică Redis, MongoDB, Memcache

și ElasticSearch. Potrivit aceluiași studiu, majoritatea acestor produse noi „*nu sunt menite să fie expuse internetului. […] Setările implicite ale acestor tehnologii tind să nu aibă nicio configurație pentru autentificare, criptare, autorizare sau orice alt tip de controale de securitate pe care le considerăm de la sine înțeles. Unele dintre ele nici măcar nu au un control de acces încorporat.*”

Mai mult, în trecut, au fost raportate incidente de partajare inadecvată a fișierelor care conțin posibile informații sensibile și confidențiale, care au afectat chiar și servicii online foarte populare precum Dropbox 10. Acest lucru este confirmat și de multe sondaje 11.

Activele vizate de aceste amenințări includ grupul de active „**Date**” și activul „**Aplicații și servicii back-end**” (cum ar fi, de exemplu, „**Servicii de facturare**”).

Amenințare: Scurgeri de date prin aplicații web (API-uri nesigure)

Diverse surse susțin că Big Data este adesea construită cu puțină securitate 12 13. Noile componente software sunt de obicei furnizate cu autorizare la nivel de serviciu, dar puține utilități sunt disponibile pentru a proteja caracteristicile de bază și interfețele de aplicație (API). Deoarece aplicațiile Big Data sunt construite pe modele de servicii web, API-urile pot fi vulnerabile la atacuri binecunoscute, cum ar fi lista Top Ten Open Web Application Security Project (OWASP) 14, cu puține facilități pentru contracararea amenințărilor web comune.

Vânzătorul de software de securitate Computer Associate (CA) 15 și alte surse 16 raportează încălcări ale datelor, din cauza API-urilor nesigure, în multe industrii, în special în rețelele sociale, în serviciile mobile de partajare a fotografiilor și video, precum Facebook, Yahoo și Snapchat.

De exemplu, o amenințare a acestei categorii poate consta în atacuri prin injectare la tehnologiile Web semantic prin injectarea codului SPARQL 17. Vulnerabilitățile de securitate sunt destul de comune în noile limbaje Big Data, cum ar fi SPARQL, RDQL (ambele sunt limbaje de interogare doar pentru citire) și SPARUL (sau SPARQL/Update, care are capabilități de modificare). Utilizarea acestor noi limbaje de interogare introduce vulnerabilități deja găsite la o utilizare proastă a limbajelor de interogare de stil vechi, deoarece atacurile precum injectarea SQL, LDAP și XPath sunt deja bine cunoscute și încă periculoase 18. Bibliotecile acestor noi limbaje oferă instrumente pentru a valida intrarea utilizatorului și pentru a minimiza riscul. Cu toate acestea, „*bibliotecile principale de limbaj de interogare ontologic încă nu oferă niciun mecanism pentru a evita injectarea de cod*” și fără aceste mecanisme, arsenalul atacatorilor ar putea fi îmbunătățit cu injecții SPARQL, RDQL și SPARQL 19. Alte noi produse software Big Data, cum

ar fi Hive, MongoDB și CouchDB, suferă, de asemenea, de amenințări tradiționale, cum ar fi execuția de cod și injecția SQL de la distanță 20.

Activele vizate de aceste amenințări aparțin grupului „**Date**” și tipului de active „**Modele de infrastructură de stocare**” (cum ar fi „**Sisteme de management al bazelor de date (DBS)**” și „**Instrumente web semantic**”).

Amenințare: Proiectare și planificare inadecvate sau adaptare incorectă

Tehnicile de îmbunătățire a performanței analizei Big Data și fuziunea surselor de date eterogene cresc redundanța ascunsă a reprezentării datelor, generând copii prost protejate. Acest lucru provoacă tehnicile tradiționale de protejare a confidențialității 21 și trebuie luat în considerare efectul redundanței. După cum s-a menționat deja, redundanța Big Data poate fi văzută ca o tehnică de atenuare a amenințărilor pentru atacuri fizice, dezastre și întreruperi 22, totuși, în unele cazuri, semnalează o slăbiciune a sistemului, fiind un stimulent de risc pentru scurgerile de Big Data. Cu alte cuvinte, dacă stocarea noastră Big Data replică înregistrările de date de zece ori și distribuie copiile la zece noduri de stocare dintr-un motiv oarecare (de exemplu, pentru a accelera conducta de analiză), cele zece noduri pot ajunge la niveluri diferite de robustețe a securității (de ex. , diferite versiuni de software de securitate) și acest lucru va crește probabilitatea dezvăluirii datelor și a scurgerilor de date. Aceasta poate fi considerată o slăbiciune specifică a designurilor Big Data.

Pe de altă parte, putem observa că chiar și redundanța și replicarea, care sunt caracteristici necesare pentru a îmbunătăți funcționalitatea Big Data, nu sunt întotdeauna sigure împotriva pierderii datelor. De exemplu, Hadoop, binecunoscutul cadru pentru procesarea Big Data, replică datele de trei ori în mod implicit, deoarece acest lucru protejează împotriva defecțiunilor inevitabile ale hardware-ului de bază. Cu toate acestea, o aplicație coruptă ar putea distruge toate replicările de date 23. De asemenea, studii recente au prezentat ideea că redundanța Hadoop ar putea fi chiar un stimulent neliniar de risc pentru scurgerile de megadate (big data) 24.

Chiar și designul sistemului de fișiere distribuit Hadoop (HDFS) semnalează probleme așa cum este raportat de literatură 25. HDFS stă la baza multor sisteme de stocare pe scară largă a Big Data și este folosit de rețelele sociale. Clienții HDFS efectuează operațiuni de metadate ale sistemului de fișiere printr-un singur server cunoscut sub numele de Namenode și trimit și recuperează datele sistemului de fișiere prin comunicare cu un grup de noduri. Pierderea unui singur nod nu ar trebui să fie niciodată fatală, dar pierderea Namenode-ului nu poate fi tolerată 26.

Rețelele sociale mari, precum Facebook, au suferit această problemă și au luat contramăsuri împotriva amenințării 27 (Hadoop instalat la Facebook include unul dintre cele mai mari clustere HDFS unice, mai mult de 100 PB de spațiu pe disc fizic într-un singur sistem de fișiere HDFS).

O altă amenințare legată de design este lipsa de scalabilitate a unor instrumente. De exemplu, NIST raportează că tehnicile originale de gestionare a drepturilor digitale (DRM) nu au fost construite pentru a fi la scară și pentru a satisface cerințele pentru utilizarea prognozată a datelor și „DRM-ul poate eșua să funcționeze în medii cu caracteristici Big Data – în special viteza și volumul agregat" 28 29.

Activele care sunt vizate de aceste amenințări aparțin grupurilor de active „**Date**" și „**Analitica Big Data**" și tipurilor de active „**Software**", „**Modele de infrastructură de calcul**" și „**Modele de infrastructură de stocare**".

Grupul de amenințări: ascultarea clandestină, interceptarea și hijacking

Acest grup include amenințările care se bazează pe alterarea/manipularea comunicațiilor dintre două părți. Aceste atacuri nu necesită instalarea de instrumente sau software suplimentare pe infrastructura victimelor.

Amenințare: Interceptarea informațiilor

O problemă comună care afectează orice infrastructură TIC este atunci când infractorii pot intercepta comunicațiile între noduri țintind legăturile de comunicație. Diverse surse susțin că comunicarea între noduri cu noile instrumente Big Data este adesea nesecurizată (30), că nu este dificil să deturnezi o sesiune de utilizator sau să obții acces neautorizat la servicii din rețelele sociale precum Facebook și Twitter (31) și că există dovezi ale unor defecte în protocoalele de comunicare. (32)

Distribuțiile de software Big Data (de exemplu Hadoop, Cassandra, MongoDB (33), Couchbase) rareori au protocoalele care asigură confidențialitatea și integritatea datelor între aplicațiile care comunică (de exemplu, TLS și SSL) activate implicit sau configurate corect (de exemplu, schimbarea parolelor implicite).

Activele vizate de această amenințare aparțin grupurilor de active „**Date**" și „**Roluri**" și activului „**Aplicații și servicii back-end**".

Grupul de amenințări: Activități/abuzuri nefaste

Acest grup include amenințările care provin din activități nefaste. Spre deosebire de grupul anterior, aceste amenințări impun (deseori) atacatorului să efectueze unele acțiuni care modifică infrastructura TIC a victimelor; de obicei cu utilizarea unor instrumente și software specifice.

Amenințare: Fraudarea identității

Sistemele Big Data stochează și gestionează acreditările pentru accesarea datelor personale și a conturilor financiare cu informații precum numerele cardurilor de credit și detaliile de plată și facturare, care sunt ținte pentru criminalii cibernetici. Sistemele Big Data stochează, de asemenea, date de profilare care pot descrie comportamentul utilizatorului, preferințele, obiceiurile, călătoriile, consumul de media la un grad ridicat de detaliu, și pot ajuta atacatorii în forme mai elaborate de fraudă prin uzurparea identității, creând oportunități mari pentru hoții de identitate (34).

Deoarece majoritatea sistemelor Big Data sunt construite peste infrastructura cloud, o amenințare la adresa identității utilizatorilor este, de exemplu, atunci când se pierde controlul unei interfețe de sistem, fie într-un sistem Big Data bazat pe un cloud public mare, fie într-un cloud privat larg utilizat (35). Un atac reușit asupra unei console oferă atacatorului putere completă asupra contului victimei, inclusiv asupra tuturor datelor stocate. Interfețele de control ar putea fi inițial compromise prin împachetarea semnăturii noi și tehnici avansate XSS, apoi escaladarea privilegiilor poate duce la fraudă de identitate (36). În timp ce în sistemele informaționale tradiționale pierderea controlului unei interfețe de consolă ar putea cauza scurgeri limitate de informații, în Big Data efectul este amplificat și impactul este mai sever.

Ingineria socială nu este o problemă nouă, dar pe măsură ce rețelele sociale devin importante atât pentru utilizatorii casnici cât și pentru companii, atacurile implică adesea ingineria socială. Atacatorii au abuzat de rețelele sociale de când au apărut pentru prima dată pe internet. De exemplu, vulnerabilitățile XSS de pe Twitter au fost folosite pentru a trimite tweet-uri rău intenționate și false, în timp ce malware-ul de internet a apărut pe Facebook ca mijloc de promovare a profilurilor rău intenționate (37).

Activele vizate de aceste amenințări sunt „Informații personale de identificare”, „Aplicații și servicii back-end” (cum ar fi, de exemplu, „Servicii de facturare”) și „Servere”.

Grupul de amenințare: Legale

Acest grup include amenințări datorate implicațiilor legale ale unui sistem Big Data, cum ar fi încălcarea legilor sau reglementărilor, încălcarea legislației, nerespectarea cerințelor contractuale, utilizarea neautorizată a resurselor de proprietate intelectuală, abuzul de date cu caracter personal, necesitatea supunerii hotărârilor judecătorești și de tribunal.

Amenințare: Încălcarea legilor sau reglementărilor / Încălcarea legislației / Abuzul de date cu caracter personal

Stocarea datelor în Uniunea Europeană intră sub incidența directivei privind protecția datelor: organizațiile sunt obligate i) să adere la această lege de conformitate pe toată durata de viață a datelor, ii) să rămână responsabile pentru datele cu caracter personal ale clienților și angajaților lor și iii) să garanteze securitatea chiar și atunci când o terță parte, cum ar fi un furnizor de cloud, prelucrează datele în numele lor.

În modelul tradițional centrat pe date, datele sunt stocate la nivel local și fiecare organizație are control asupra informațiilor. În schimb, în Big Data, apare o preocupare reală cu privire la securitatea acestei cantități masive de informații digitale și protecția infrastructurii critice care o susține, așa cum demonstrează o vastă literatură despre riscurile de confidențialitate (38) (39) (40) (41).

De asemenea, trebuie menționat că UE are reglementări mai stricte în ceea ce privește colectarea datelor cu caracter personal decât alte țări, dar uneori multinaționalele care operează în UE au sediul în Statele Unite. În acest context, cele mai importante probleme de confidențialitate sunt modul de protejare a confidențialității individuale atunci când datele sunt stocate pe mai multe site-uri și cât de eficientă este protecția la.

Big Data ridică, de asemenea, potențiala problemă a rezidenței datelor (42). Datele, atunci când sunt stocate în stocarea cloud a furnizorilor care oferă soluții de stocare multinaționale, pot intra în jurisdicții legale diferite. Un exemplu adus de NIST Big Data Public Working Group se referă la custodia datelor farmaceutice dincolo de dispozițiile de testare, care este neclară, mai ales după fuzionarea sau dizolvarea firmelor (43).

Activele vizate de această amenințare includ grupuri de active „**Date**” (în special „**date de înregistrare de identificare**”) și „**Roluri**”.

Grup de amenințări: Amenințări organizaționale

Acest grup include amenințările care țin de sfera organizațională.

Amenințare: Lipsa de calificare

Analiza seturilor mari de date poate susține noi valuri de creștere a productivității și inovației și poate debloca o valoare semnificativă. Cu toate acestea, companiile și factorii de decizie politică trebuie să abordeze obstacole semnificative, cum ar fi, de exemplu, o posibilă lipsă de oameni de știință de date și manageri calificați (44).

Activul vizat de această amenințare este grupul de active „**Roluri**”.

Referințe

(1) Smart Grid Threat Landscape, Threat Landscape and Good Practice Guide for Internet Infrastructure, Threat Landscape and Good Practice Guide for Smart Home and Converged Media

(2) ENISA, „Threat Landscape and Good Practice Guide for Internet Infrastructure”, ianuarie 2015

(3) Trebuie să menționăm că unele instalații Big Data de bază, de testare sau cu scop special, ar putea să nu se bazeze, total sau parțial, pe redundanță [cloud]. De exemplu, Big Data instalat într-un mediu cloud privat, fără servicii de replicare activate, nu are opțiune de recuperare în caz de dezastru natural. Fiind limitate la un mediu privat de cloud local, aceste instalații ar putea fi, în principiu, supuse unor atacuri fizice și dezastre naturale și de mediu, cum ar fi cutremure, inundații, alunecări de teren, tsunami, incendii, poluare, praf, tunete și alte evenimente majore. pentru mediu. Cu toate acestea, activarea serviciilor private de replicare în cloud între diferite locații fizice diminuează acest risc.

(4) A se vedea glosarul din https://www.enisa.europa.eu/activities/risk-management/current-risk/risk-management-inventory/glossary, accesat în decembrie 2015.

(5) E. Damiani, „Toward Big Data Leak Analysis”. Proceedings of Privacy and Security of Big Data Workshop (PSBD 2015), IEEE Big Data Conference, San Jose, CA, 1-3 noiembrie 2015

(6) A se vedea modelul ISO 15408.

(7) A se vedea taxonomia erorii umane în sistemele informaționale în Im și Richard L. Baskerville (Georgia State University), „*A Longitudinal Study of Information System Threat Categories: The Enduring Problem of Human Error*”. ACM SIGMIS (2005).

(8) Conform „*IBM Security Services 2014 Cyber Security Intelligence Index*” peste 95% din toate incidentele investigate recunosc „eroarea umană” și cea mai răspândită eroare umană este„Dublu clic” pe un atașament infectat sau pe o adresă URL nesigură.

(9) BinaryEdge, o firmă de inginerie de securitate cu sediul în Elveția, a investigat patru tehnologii obișnuite de Big Data, cum ar fi Redis, MongoDB, Memcache și Elasticsearch și a găsit diverse probleme de configurare. De exemplu, compania a descoperit că zeci de mii de instanțe de baze de date NoSQL erau accesibile fără a fi necesară autentificarea. A se vedea http://blog.binaryedge.io/2015/08/10/data-technologies-and-security-part-1/, accesat în decembrie 2015.

(10) Techcrunch, un editor online popular de știri din industria tehnologiei, a raportat că la Dropbox, pentru o perioadă scurtă de timp, serviciul a permis utilizatorilor să se conecteze la conturi folosind orice parolă. Cu alte cuvinte, oamenii se puteau conecta la contul cuiva doar introducând adresa lor de e-mail. Consultați http://techcrunch.com/2011/06/20/dropbox-security-bug-made-passwords-optional-for-four-hours/, accesat în decembrie 2015.

(11) Marea majoritate (mai mult de 80%) a participanților la raportul *EMA Research* din 2015 privind securitatea colaborării cu fișierele, sponsorizat de FinalCode, au recunoscut că au existat incidente de scurgere de date în organizațiile lor. Consultați http://www.finalcode.com/en/how-it-works/resources/ema-report/, accesat în decembrie 2015.

(12) *Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments*, publicat de compania de securitate Securosis, în octombrie 2012, https://securosis.com/assets/library/reports/SecuringBigData_FINAL.pdf, accesat în decembrie 2015.

(13) Eduardo B. Fernandez (Departamentul de Calcul și Inginerie Electronică și Știința Calculatoarelor, Florida Atlantic University), „Security in Data Intensive Computing Systems in Handbook of Data Intensive Computing". *Springer* (2011), http://link.springer.com/chapter/10.1007/978-1-4614-1415-5_16, accesat în decembrie 2015.

(14) Multe expuneri comune la vulnerabilități pentru componentele Big Data, cum ar fi Hadoop, sunt raportate pe site-uri web specializate, a se vedea, de exemplu, https://cve.mitre.org și https://www.cvedetails.com, accesate în decembrie 2015.

(15) Jaime Ryan (CA, Director Sr.) și Tyson Whitten (CA, Director of API Management) în prezentarea și webinarul „*Takeaways from API Security Breaches*" (2015) au raportat încălcări, din cauza API-urilor nesecurizate, pentru Yahoo, Snapchat și alte companii , consultați http://transform.ca.com/API-security-breaches.html?source=AAblog, accesat în decembrie 2015.

(16) A se vedea problemele de securitate pentru biblioteca Graph Facebook API raportate de blogul tehnic *Websegura*, http://www.websegura.net/advisories/facebook-rfd-and-open-file-upload/, accesat în decembrie 2015.

(17) A se vedea http://www.morelab.deusto.es/code_injection/ și următoarea publicație: Pablo Orduña, Aitor Almeida, Unai Aguilera, Xabier Laiseca, Diego López-de-Ipiña, Aitor Gómez-Goiri, „Identifying'Identifying Security Issues in the Semantic Web: Injection attacks in the Semantic Query Languages", [*VI Jornadas Científico-Técnicas en Servicios Web y SOA* (JSWEB 2010p.)], Valencia, Spania. Septembrie 2010, p. 43 - 50. ISBN: 978-84-92812-59-2.

(18) În octombrie 2015, probabil, o injecție SQL a fost folosită pentru a ataca serverele companiei britanice de telecomunicații Talk Talk's, punând în pericol detaliile personale a până la patru milioane de clienți. A se vedea http://www.mobilenewscwp.co.uk/2015/10/23/talktalk-hacking-scandal-expert-reaction/, accesat în decembrie 2015.

(19) Ben Mustapha et al., „Enhancing semantic search using case-based modular ontology". în *Proceeding of the 2010 ACM Symposium on Applied Computing*.

(20) De exemplu, versiunea 2.0 Hive suferă de criptare încrucișată, execuție de cod și vulnerabilități de injectare SQL la distanță, consultați https://packetstormsecurity.com/files/132136/Hive-2.0-RC2-XSS-Code-Execution-SQL-Injection. html. MongoDB suferă atacuri prin injecție, vezi https://www.idontplaydarts.com/2011/02/mongodb-null-byte-injection-attacks/. Vedeți și alte amenințări specifice furnizorului în prezentare https://www.defcon.org/images/defcon-21/dc-21-presentations/Chow/DEFCON-21-Chow-Abusing-NoSQL-Databases.pdf, accesat în decembrie 2015.

(21) E. Damiani, „*Toward Big Data Risk Analysis*”, Discurs principal la cel de-al 2-lea Atelier internațional de confidențialitate și securitate a datelor mari (PSBD 2015)

(22) Atacurile fizice, dezastrele și, respectiv, întreruperile sunt descrise ca grup de amenințări în taxonomia generică a amenințărilor ENISA.

(23) A se vedea http://www.smartdatacollective.com/michelenemschoff/193731/how-your-hadoop-distribution-could-lose-your-data-forever, accesat în decembrie 2015.

(24) E. Damiani, „*Toward Big Data Leak Analysis*”, Proceedings of the Privacy and Security of Big Data Workshop (PSBD 2015), IEEE Big Data Conference, San Jose, CA, 1-3 noiembrie 2015

(25) Aditham, Ranganathan (Departamentul de Informatică și Inginerie, Universitatea din Florida de Sud, Tampa, SUA), „*A Novel Framework for Mitigating Insider Attacks in Big Data Systems*”. 2015 IEEE International Conference on Big Data

(26) Toate operațiunile cu metadate trec prin Namenode. Dacă Namenode nu este disponibil, niciun client nu poate citi sau scrie pe HDFS, iar utilizatorii și aplicațiile care depind de HDFS nu vor putea funcționa corect. Versiunile recente de Hadoop au introdus și alte componente pentru gestionarea resurselor pentru a rezolva această problemă.

(27) A se vedea „*Notes by Facebook engineering*” în https://www.facebook.com/notes/facebook-engineering/under-the-hood-hadoop-distributed-filesystem-reliability-with-namenode-and-avata/10150888759153920. Facebook a contribuit la o soluție funcțională pentru a rezolva deficiențele arhitecturale ale failoverului unic Namenode, numit Avatarnode. Acesta este un modul cu sursă deschisă care oferă failover și failback la cald și este acum în producție la Facebook, rulând cel mai mare cluster Hadoop Data Warehouse (100 PB spațiu pe disc fizic într-un singur sistem de fișiere HDFS).

(28) A se vedea *NIST Special Publication* 1500-4. Use case: consumer digital media (exemple: Netflix, iTunes și altele).

(29) Xiao Zhang, „*A Survey of Digital Rights Management Technologies*”, vezi http://www.cse.wustl.edu/~jain/cse571-11/ftp/drm.pdf, accesat în decembrie 2015.

(30) *Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments*, publicată de compania de securitate Securosis, L.L.C., 12 octombrie 2012.

(31) A se vedea, de exemplu, „*How to prevent a session hijacking attack*” in Facebook și Twitter, http://searchmidmarketsecurity.techtarget.com/tip/Defending-against-Firesheep-How-to-prevent-a-session-hijacking-attack, accesat în decembrie 2015.

(32) A se vedea, de exemplu, un atac împotriva confidențialității datelor aflate în tranzit prin rețelele nesigure în http://www.isg.rhul.ac.uk/tls/Lucky13.html, accesat în decembrie 2015.

(33) A se vedea, de exemplu, documentația MongoDB despre aceasta și greșelile care pot compromite baza de date (erori de configurare TTL și altele), http://blog.mongodb.org/post/87691901392/mongodb-security-part-ii-10-mistakes- that-can, accesat în decembrie 2015.

(34) Big data creează mari oportunități pentru hoții de identitate: a se vedea http://www.c4isrnet.com/story/military-tech/it/2015/01/19/big-data-identity-theft/22004695/, accesat în decembrie 2015.

(35) A se vedea: J. Somorovsky et al., „All your clouds belong to us: security analysis of cloud management interfaces”, în *Proceedings of the 3rd ACM workshop on Cloud computing security workshop* (http://dl.acm.org/citation.cfm?id=2046664) pentru atacuri la Amazon și Eucalyptus. Lucrarea oferă o analiză de securitate a interfețelor de control ale serviciilor Cloud publice mari (Amazon EC2 și S3) și software-ului Cloud privat (Eucalyptus).

(36) A se vedea http://www.zdnet.com/article/us-cert-warns-of-guest-to-host-vm-escape-vulnerability/. Articolul descrie o vulnerabilitate, care afectează sistemele de operare pe 64 de biți și software-ul de virtualizare care rulează pe hardware CPU Intel și expune utilizatorii la atacuri locale de escaladare a privilegiilor sau la o evadare de la o mașină virtuală de la oaspete la gazdă.

(37) A se vedea *Nine Threats Targeting Facebook Users* în http://www.itbusinessedge.com/slideshows/show.aspx?c=90875, accesat în decembrie 2015.

(38) Venkat N. Gudivada (Universitatea Marshall), Ricardo Baeza-Yates (Laboratoarele Yahoo), Vijay V. Raghavan (Universitatea din Louisiana), „Big Data: Promises and Problems", Numărul nr.03, 2015, publicat de *IEEE Computer Societate*. Consultați http://www.computer.org/csdl/mags/co/2015/03/mco2015030020.html. Cartea afirmă că „veridicitatea – datorită prelucrării intermediare, diversității între sursele de date și în evoluția datelor ridică îngrijorări cu privire la securitate, confidențialitate, încredere și responsabilitate, creând necesitatea de a verifica proveniența securizată a datelor".

(39) *White House Big Data Report*, publicat la 1 mai 2014. A se vedea https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf, accesat în decembrie 2015.

(40) A se vedea diverse exemple date de Raymond Chi-Wing Wong, „Big Data Privacy", în *Journal of Information Technology & Software Engineering, Department of Computer Science and Engineering*, Hong Kong University of Science and Technology, China, http://www. omicsgroup.org/journals/big-data-privacy-2165-7866.1000e114.php?aid=10289. Articolul citează diferite cazuri de utilizare: (i) datele persoane ale unui set de date medicale au fost identificate din cauza protecției insuficiente a confidențialității, (ii) seturi de date, inclusiv jurnalele de căutare, au fost eliberate de un furnizor de internet american, dar a fost posibil să se identifice o persoană folosind mai multe persoane- interogări specifice, (iii) un serviciu popular de închiriere de filme online cu un sistem de recomandare, filme propuse clienților săi pe baza preferințelor lor de film anticipate, cu toate acestea, aproape toți abonații ar putea fi identificați în mod unic, (iv) mulți clienți de telefonie mobilă folosind bazate pe locație serviciile (LBS) au avut serioase preocupări legate de confidențialitate cu privire la dezvăluirea locațiilor lor împreună cu informațiile lor personale.

(41) S. De Capitani di Vimercati, S. Foresti, P. Samarati, „Managing and Accessing Data in the Cloud: Privacy Risks and Approaches", în *Proc. of the 7th International Conference on Risks and Security of Internet and Systems (CRiSIS 2012)*, Cork, Irlanda.

(42) Vezi *Storing Data In The Cloud Raises Compliance Challenges* în http://www.forbes.com/sites/ciocentral/2012/01/19/storing-data-in-the-cloud-raises-compliance-challenges/, accesat în decembrie 2015.

(43) A se vedea *NIST Special Publication 1500-4. Use case: Pharmaceutical clinical trial data sharing.*

(44) A se vedea, de exemplu, rapoarte de la *McKinsey* http://www.mckinsey.com/features/big_data și de la *Financial Times* http://www.ft.com/cms/s/0/953ff95a-6045-11e4-88d1-00144feabdc0 .html#axzz3ntU3lM00, accesat în decembrie 2015.

Sursa: Sfetcu, Nicolae (2022). *Big Data: Modele de afaceri - Securitatea megadatelor*, MultiMedia Publishing, ISBN 978-606-033-655-6, https://www.telework.ro/ro/e-books/big-data-modele-de-afaceri-securitatea-megadatelor/